BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37082
doi: 10.3897/biss.3.37082

OPEN ACCESS

Conference Abstract

Extracting Data from Legacy Taxonomic Literature: Applications for planning field work

Francisco Andres Rivera-Quiroz[‡], Jeremy A. Miller[‡,§]

‡ Naturalis Biodiversity Center, Leiden, Netherlands
§ Plazi, Bern, Switzerland

Corresponding author: Francisco Andres Rivera-Quiroz (andres.riveraquiroz@naturalis.nl)

Received: 11 Jun 2019 | Published: 18 Jun 2019

Citation: Rivera-Quiroz F, Miller J (2019) Extracting Data from Legacy Taxonomic Literature: Applications for planning field work. Biodiversity Information Science and Standards 3: e37082. https://doi.org/10.3897/biss.3.37082

Abstract

Traditional taxonomic publications have served as a biological data repository accumulating vast amounts of data on species diversity, geographical and temporal distributions, ecological interactions, taxonomic relations, among many other types of information. However, the fragmented nature of taxonomic literature has made this data difficult to access and use to its full potential. Current anthropogenic impact on biodiversity demands faster knowledge generation, but also making better use of what we already have. This could help us make better-informed decisions about conservation and resources management.

In past years, several efforts have been made to make taxonomic literature more mobilized and accessible. These include online publications, open access journals, the digitization of old paper literature and improved availability through online specialized repositories such as the Biodiversity Heritage Library (BHL) and the World Spider Catalog (WSC), among others. Although easy to share, PDF publications still have most of their biodiversity data embedded in strings of text making them less dynamic and more difficult or impossible to read and analyze without a human interpreter. Recently developed tools as GoldenGATE-Imagine (GGI) allow transforming PDFs in XML files that extract and categorize taxonomically relevant data. These data can then be aggregated in databases such as Plazi TreatmentBank, where it can be re-explored, queried and analyzed.

Here we combined several of these cybertaxonomic tools to test the data extraction process for one potential application: the design and planning of an expedition to collect fresh material in the field. We targeted the ground spider *Teutamus politus* and other related species from the *Teutamus* group (TG) (Araneae; Liocranidae). These spiders are known from South East Asia and have been cataloged in the family Liocranidae; however, their relations, biology and evolution are still poorly understood. We marked-up 56 publications that contained taxonomic treatments with specimen records for the Liocranidae. Of these publications, 20 contained information on members of the TG. Geographical distributions and occurrences of 90 TG species were analyzed based on 1,309 specimen records. These data were used to design our field collection in a way that allowed us to optimize the collection of adult specimens of our target taxa.

The TG genera were most common in Indonesia, Thailand and Malaysia. From these, Thailand was the second richest but had the most records of *T. politus*. Seasonal distribution of TG specimens in Thailand suggested June and July as the best time for collecting adults. Based on these analyses, we decided to sample from mid-July to mid-August 2018 in the three Thai provinces that combined most records of TG species and *T. politus*.

Relying on the results of our literature analyses and using standard collection methods for ground spiders, we captured at least one specimen of every TG genus reported for Thailand. Our one-month expedition captured 231 TG spiders; from these, *T. politus* was the most abundant species with 188 specimens (95 adults). By comparison, a total of 196 specimens of the TG and 66 of *T. politus* had been reported for the same provinces in the last 40 years. Our sampling greatly increased the number of available specimens, especially for the genera *Teutamus* and *Oedignatha*. Also, we extended the known distribution of *Oedignatha* and *Sesieutes* within Thailand.

These results illustrate the relevance of making biodiversity data contained within taxonomic treatments accessible and reusable. It also exemplifies one potential use of taxonomic legacy data: to more efficiently use existing biodiversity data to fill knowledge gaps. A similar approach can be used to study neglected or interesting taxa and geographic areas, generating a better biodiversity documentation that could aid in decision making, management and conservation.

Keywords

Araneae, Liocranidae, Teutamus, cybertaxonomy, Plazi, PDF, XML, GoldenGATE-Imagine

Presenting author

Francisco Andres Rivera-Quiroz

Presented at

Biodiversity_Next 2019